

文章编号: 2096-1618(2018)02-0143-06

基于时空兴趣点的人体动作识别

陈胜娣, 何冰倩, 陈思宇, 刘基缘

(成都信息工程大学计算机学院, 四川 成都 610225)

摘要: 人体动作识别在计算机视觉研究和模式识别领域中逐渐成为一个研究热点。提出一种基于 Harris-Laplace 时空兴趣点结合 3D-SIFT 描述子, 通过 Bag-of-feature 构建词袋的方法, 并应于用人体动作识别。针对传统 Harris 算法提取出的兴趣点冗余, 所以采用 Harris-Laplace 算法提取时空兴趣点。3D-SIFT 描述子能更好地描述视频序列的本质特征, 并且比传统的描述子更有效, Bag-of-feature 词袋法表征特征, 采用改进的 K 均值 (K -Means) 聚类算法进行聚类, 最后采用多分类支持向量机 (SVM) 进行一对一、一对多的分类策略并进行比较。在 KTH 公开运动数据集上进行实验测试, 实验结果证明提出的人体动作识别方法的有效性和鲁棒性。

关键词: 计算机应用技术; 图像图形处理; 时空兴趣点; 动作识别; Harris-Laplace; 3D-SIFT; 特征提取

中图分类号: TP391.41

文献标志码: A

doi: 10.16836/j.cnki.jcui.2018.02.007

0 引言

人体动作识别在机器学习, 计算机视觉和人机交互等领域中逐渐成为一个研究热点。被广泛应用于公共场所的视频监控, 手语分析, 人机交互, 基于内容的图像存储和检索, 虚拟现实^[1], 对患有帕金森的人的健康监测和运动监护^[2], 人体动作能量消耗评估等方面, 同时在许多的工业以及军事领域都有着广泛的应用前景^[3-4]。人体动作识别是通过视觉信息处理和分析技术对来自摄像机的视频数据中的人体动作所传递的信息进行语义描述^[5]。相对于在图像序列中进行人体检测、跟踪, 人体动作识别属于更高层次的视觉任务, 是当前视觉研究领域备受关注和最具挑战性的研究方向之一^[6]。动作识别主要有特征的提取和动作分类^[7]两个步骤; 而更鲁棒的特征表征方法, 是识别成功的关键因素。目前的动作识别还面临着很多的挑战, 由于人体运动的不确定性姿态的随意性、光照、遮挡等因素, 都对特征点的提取造成一定的干扰作用, 对准确分类识别带来了很大的困难。

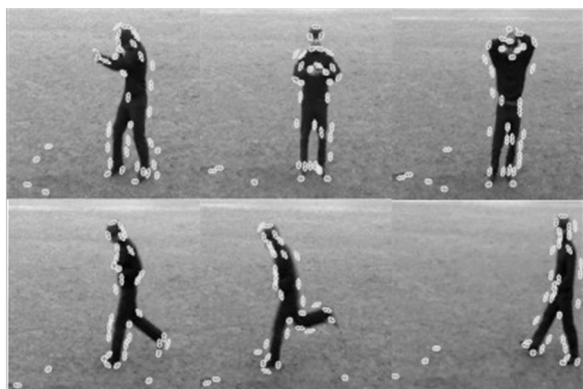
目前在动作识别领域, 做了很多的研究工作。Cherian 等^[8]通过提取 IDT 特征, 利用卷积神经网络进行训练分类识别。该方法虽然能得到比较好的识别结果, 但是计算量较大, 对硬件的需求较高, 并且需要的训练数据集也要求比较大。Laptev 等^[9]提出将 2D-Harris 角点扩展为 3D-Harris, 使得到的时空特征点邻域的像素值能在时间和空间维度上都有显著性变化。

Park 等^[10]提出一种基于贝叶斯网络的人体姿态估计放到对每个个体的动作进行建模, 最后通过决策树的方法对动作进行分类。丁松涛等^[11]提出改进的 Harris-Laplace 算法, 通过对背景点抑制和时间抑制来提高兴趣点的提取精度。Guan 等^[12]采用隐马尔科夫 (HMM) 方法进行动作识别。这些方法都在实验中取得一定的成功, 但是计算量大, 计算复杂度影响了分类的效率。Schuldt 等^[13]提出将时空兴趣点作为特征描述子, 利用 K -Means 进行聚类, 最后通过 SVM 进行动作分类识别。在文献 [13-14] 的基础上, 提出利用 Harris-Laplace 时空兴趣点提取 3D-SIFT 描述子, 通过构建词袋进行改进的 K -Means 聚类^[15], 并用 SVM 进行动作分类。该方法在 KTH^[16]数据集上进行验证, 实验结果表明, 提出的方法的有效性。

1 动作识别系统

相对于全局特征 (光流, 形状, 运动轨迹) 只有在固定的摄像机和简单的背景以及分辨率比较高的特定场合下识别率才比较理想, 如果出现部分遮挡, 噪声光线等干扰的情况下识别效果欠佳。基于局部运动的时空兴趣点就很好地弥补了这方面的不足。时空兴趣点能很好地反应视频中图像变换显著的区域, 具有很好的特征描述和区分辨别的能力^[17]。相对于传统的 Harris 角点, 提取出的兴趣点比较冗余并且精度不高。采用的 Harris-Laplace 算法在角点检测的过程中包含对图像的空间信息的处理和对角点进行过滤的操作, 这能有效减少无用背景信息点的产生, 为后期对检测到的角点进行特征提取和动作识别提供更为有效和更

为准确的数据支撑。图1是传统的 Harris 角点检测和应用的 Harris-Laplace 算法提取的时空兴趣点的分布对比图。从图1可以看出,传统的 Harris 角点检测算法检测到大量无用的兴趣点,而应用 Harris-Laplace 算法检测到的兴趣点更为精确,这将提高后期图像特征的计算处理速度。例如图1(a)中的第一个个体的“拳击”动作,传统的 Harris 算法检测到整个个体周身大量无关的兴趣点和一些不具运动信息的背景点,而从图1(b)中第一个个体的检测结果可以看出,Harris-Laplace 算法检测到的兴趣点主要集中在个体的手部拳击运动区域。Harris-Laplace 检测关键点分为2个步骤:多尺度 Harris 角点检测^[18-19],然后利用 LoG(边缘检测)算子对角点进行过滤^[20]。



(a) 传统算法提取时空兴趣点



(b) 文中算法提取时空兴趣点

图1 KTH 数据集多尺度时空兴趣点实验结果对比

1.1 Harris-Laplace 时空兴趣点检测

Harris 角点对光线和对比度的变化具有鲁棒性,但是在尺度变化上比较敏感,Harris-Laplace 算法检测到的角点带有尺度信息。通过将视频序列与变换高斯核函数滤波器进行卷积操作得到不同的尺度空间的图像序列^[11]。图像尺度空间的计算如下:

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (2)$$

式(1)是将输入图像 $I(x, y)$ 与尺度因子为 σ 的高斯核函数 $G(x, y, \sigma)$ 进行卷积操作得到图像尺度空间 $L(x, y, \sigma)$ 。其中高斯核函数的计算如式(2)所示。

然后,计算每一层尺度空间图像的 Harris-Laplace 自相关矩阵:

$$M = u(x, y, \sigma_l, \sigma_d) = \sigma_d^2 g(\sigma_l) \otimes \begin{pmatrix} L_x^2(x, \sigma_d) & L_x L_y(x, \sigma_d) \\ L_x L_y(x, \sigma_d) & L_y^2(x, \sigma_d) \end{pmatrix} \quad (3)$$

$$L_x^2 = L_x \times L_x \quad (4)$$

$$L_y^2 = L_y \times L_y \quad (5)$$

$$L_x L_y = L_x \times L_y \quad (6)$$

式(3)中, x, y 表示图像像素坐标位置, σ_l 表示积分尺度, σ_d 表示微分尺度。一般有 $\sigma_l = \alpha \sigma_d$, 在实验中取 $\alpha = 0.7$ 。 L_x, L_y 分别表示尺度空间函数 L 对 x, y 的偏导数。

接下来,根据式(3)计算得到的自相关矩阵,来计算每一层尺度 σ 上图像的各个像素点 (x, y) 的 Harris 响应值:

$$R = \det(u(x, y, \sigma_l, \sigma_d)) - \kappa \cdot \text{trace}^2(u(x, y, \sigma_l, \sigma_d)) > T \quad (7)$$

式(7)中, κ 表示一个常数,一般取值 $0.04 \sim 0.06$ (文中 $\kappa = 0.06$)。 T 为控制角点数目的阈值。其中 Harris 响应值 R 越大则表明该点越有可能是 Harris 角点。

最后根据式(7)计算得到的 Harris 响应值 R 来提取每一层尺度图像的 Harris 角点。具体的提取方法是:以输入图像的某点为中心取 3×3 的比较窗口,如果该点的 Harris 响应值 R 大于它周围的8个点的 R 值并且大于设定的阈值 T ,则该点被提取为 Harris 角点。

1.2 LoG 筛选角点

对于图像尺度空间检测出的 Harris 角点 (x, y) , 利用 LoG 算子将尺度变化敏感的角点进行过滤。LoG 算子的计算如下:

$$|LoG(x, \sigma_n)| = \sigma_n^2 |L_{xx}(x, \sigma_n) + L_{yy}(x, \sigma_n)| \quad (8)$$

式中, L_{xx} 和 L_{yy} 分别表示尺度函数对 x, y 的二阶偏导数, σ_n 为第 n 层的尺度。

1.3 3D-SIFT 描述子

对于视频帧中提取到的 Harris-Laplace 时空兴趣点的区域进行提取 3D-SIFT 描述子信息。SIFT 特征对于两幅图像之间发生平移,旋转,仿射变换等情况下都具有很强的匹配能力。3D-SIFT 描述子^[21]。是对

2D SIFT 的一个扩展。对于 2D-SIFT 和 3D-SIFT 的区别,如图 2 所示。

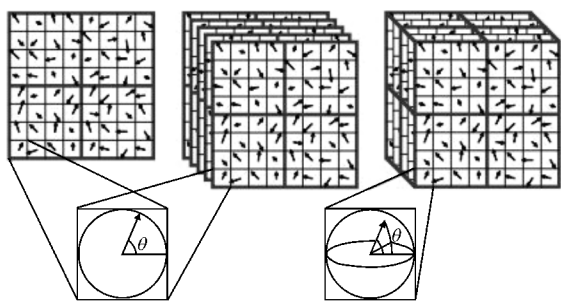


图2 2D-SIFT 描述子和 3D-SIFT 描述子

对每一层尺度空间 σ 下提取到的每一个 Harris-Laplace 角点 (x, y) 计算它的 2D-SIFT 描述子的灰度梯度幅值 m_{2D} 和梯度方向 θ 的公式如下:

$$m_{2D}(x, y) = \sqrt{H_x^2 + H_y^2} \quad (9)$$

$$\theta(x, y) = \tan^{-1}\left(\frac{H_y}{H_x}\right) \quad (10)$$

$$H_x = \sigma(x+1, y) - \sigma(x-1, y) \quad (11)$$

$$H_y = \sigma(x, y+1) - \sigma(x, y-1) \quad (12)$$

其中 $\sigma(x, y)$, H_x , H_y 分别表示角点 (x, y) 所在的空间尺度,水平方向梯度和垂直方向梯度。

3D-SIFT 描述子是在 2D-SIFT 描述子的基础上加入时间 t 作为第三个维度信息,实现从二维静态图像扩展到计算三维动态视频序列的 3D-SIFT 描述子。结合图 2 可以看出,在三维球面坐标系中,需要有两个方向角才能确定角点的梯度方向向量。3D-SIFT 描述子的灰度梯度幅值 m_{3D} 和梯度方向向量的计算如下

$$m_{3D}(x, y, t) = \sqrt{H_x^2 + H_y^2 + H_t^2} \quad (13)$$

$$\theta(x, y, t) = \tan^{-1}\left(\frac{H_y}{H_x}\right) \quad (14)$$

$$\varphi(x, y, t) = \tan^{-1}\left(\frac{H_t}{\sqrt{H_x^2 + H_y^2}}\right) \quad (15)$$

$$H_t = H(x, y, t+1) - H(x, y, t-1) \quad (16)$$

相比 2D-SIFT 描述子,3D-SIFT 描述子在计算过程中需要计算时间方向上的梯度 H_t 。式(14)中的 θ 表示向量在 xy 平面上的投影与 x 轴方向的夹角。式(15)中 φ 表示向量与 xy 平面的夹角。由于 φ 的取值范围是 $(-\frac{\pi}{2}, \frac{\pi}{2})$,所以在三维空间中,对于每一个角点的梯度方向都有唯一的一组 (θ, φ) 与之对应。

通过式(13)、(14)、(15)计算出视频序列 (x, y, t) 的每一个时空兴趣点的梯度幅值和梯度方向,最后以兴趣点为特征点中心,利用高斯权重函数在立方体视频序列中做梯度集成,得到向量化的描述子。

1.4 构建视觉词袋

通过前文的 3D-SIFT 得到特征点的描述子,改进的 K -Means 聚类算法将所有的特征描述子进行 K 聚类。 K -Means 算法^[22]是一种基于样本间相似性度量的间接聚类方法,将所有的描述子分为 K 个类簇,聚类的过程是根据使得簇内描述子相似度比较高,而类簇之间的相似度比较低的原则。传统的 K -Means 算法主要有 4 个计算步骤:

步骤 1 从数据集中随机选取 K (预先定义好的) 个样本作为初始化聚类中心 $C = \{c_1, c_2 \dots c_K\}$;

步骤 2 对于数据集中的每一个样本 f_i , 分别计算它到 K 个聚类中心的距离,然后将它分类到距离聚类中心最小距离的那个类中;

步骤 3 对于每一个类别 c_i , 重新计算它的聚类中心,也就是属于该类的所有样本的质心。计算公式如下:

$$c_i = \frac{1}{|c_i|} \sum_{f \in c_i} f \quad (17)$$

步骤 4 重复步骤 2、步骤 3,直到聚类中心的位置不再发生变化为止。

相对于传统的 K -Means 聚类算法,一开始就随机确定 K 个聚类中心,而改进的 K -Means 算法是先随机确定一个聚类中心,剩下的 $K-1$ 个聚类中心通过概率筛选进行确定,每次要确定的下一个聚类中心都是根据聚类中心相对距离越远越好的原则,这样类簇之间就比较分散,从而实类间相差较大,类内相差较小的原则,以此来提高动作的识别准确率。改进的 K -Means 算法步骤如下:

步骤 1 对于所有的特征描述子 F , 随机确定第一个聚类中心 c_1 ;

步骤 2 计算所有特征描述子与当前聚类中心之间的最短距离 $D(f)$; 再根据式(18)计算每个样本被选为下一个聚类中心的概率,最后按照轮盘赌算法选择出下一个聚类中心。

$$p(i) = \frac{D(f_i)^2}{\sum_{f \in F} D(f_i)^2} \quad (18)$$

轮盘赌选择算子^[23-25],在遗传算法中常常被选作为选择算子,根据个体的适应度按比例转为被选择概率,根据选择概率在圆盘上进行比例划分,最后转动转盘,指针停止时所指的扇区即为选中的个体。由于扇区的面积和适应度的值成正比,所以对于适应度值比较大的个体,它被选中的概率也比较大。在改进的 K -Means 算法中,轮盘赌算法选择下一个聚类中心的步骤如下:

Step 1: 每个个体的适应度为(2)中的 $D(f)$;

Step 2: 其中每个个体被选中为下一个聚类中心的概率为 $p(i)$,根据 $p(i)$ 把圆盘分成 n 个扇区,其中扇区的中心角与 $p(i)$ 的大小成正比;

Step 3: 产生一个随机数 $a \in [0,1]$,如果 a 满足式(19),则特征描述子 f_i 被选为下一个聚类中心。

$$\sum_{j=1}^{i-1} p_j a \sum_{j=1}^i p_j$$

(19)

步骤3 重复步骤2,直到最后选出 K 个聚类中心。

步骤4 后面的计算过程与传统的 K -Means 中的步骤2到步骤4相同。

在 BOF 模型中,将聚类中心称为视觉词, K 就是构成的视觉词袋的码本长度,对于所有的特征描述子,计算它们与 K 个聚类中心的距离,将其映射到距离最近的聚类中心(视觉词汇)中。这样就通过改进的 K -Means 聚类算法,构造出特征描述子码本。最后,用直方图来统计单幅图像中每一类特征描述子出现的次数,从而得到一张概率直方图。在实验中取 $K=600$ 。具体过程如图3所示。

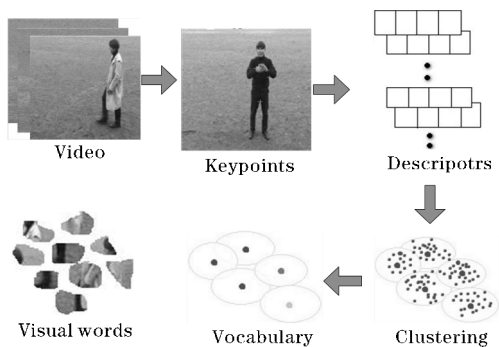


图3 构建视觉词袋的过程

1.5 支持向量机(SVM)分类

通过上面的方法得到图像的概率直方图数, $H = \{h_1, h_2, h_3, \dots, h_n\}$, h_i 是每一类特征描述子出现的概率,单幅图像的特征点的总类别数是 n . 利用 χ^2 核函数非线性支持向量机(SVM)来进行分类,对于特征的相似性,通过计算直方图 H_i, H_j 之间的距离来判断^[13]。其中,核函数的计算如下:

$$K(H_i, H_j) = \exp\left[-\frac{1}{2A} \sum_{m=1}^n \frac{(h_{im} - h_{jm})^2}{h_{im} + h_{jm}}\right]$$

(20)

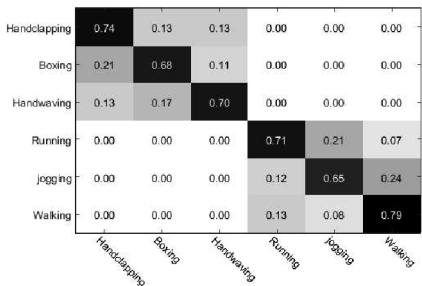
其中, $H_i = \{h_{i1}, h_{i2}, h_{i3}, \dots, h_{in}\}$,同理, $H_j = \{h_{j1}, h_{j2}, h_{j3}, \dots, h_{jn}\}$,用来表示时空中单词出现的频率直方图,而 n 表示的是时空码本的维数。

2 实验结果及分析

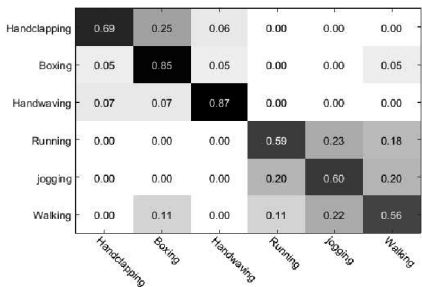
在 Windows7 平台 MATLAB 实验环境下,将提出

的方法应用在公开的动作数据集 KTH 上进行实验评估。该数据集包含 25 个人在 4 种不同场景(室外、室外尺度缩放变化、室外个体衣着变化、室内)下完成的 6 类常见动作,共有 2391 段视频序列。6 类动作分别为:鼓掌(Hand clapping)、拳击(Boxing)、挥手(Hand waving)、跑步(Running)、慢跑(jogging)、行走(walking)。在实验中,随机选取80%作为训练集,20%为测试集。分别用了一对一的 SVM 分类策略和一对多的分类策略进行比较。动作识别准确率的混淆矩阵如图4所示。从图4可以看出,不同的分类策略对动作识别的结果有一定的影响,例如相对“行走”动作,在图4(a)一对一分类策略中识别率为0.79,在图4(b)一对多分类策略中为0.56,主要被误判为“慢跑”动作。而“鼓掌”、“拳击”、“挥手”这3个动作的识别率对不同的分类策略识别效果相对稳定。

表1列出文中的动作识别方法与文献[14],在相同训练集和相同测试集上的动作识别率对比。从表1可以看出,文中的平均准确率比文献[14]平均提高了6%。



(a) 1vs. 1 策略混淆矩阵



(b) 1vs. rest 策略混淆矩阵

图4 混淆矩阵

表1 动作识别准确率比较

| 方法 | 1 vs. 1 | 1 vs. rest | 平均 |
|----------------------------|---------|------------|---------|
| Sushirdeep ^[14] | 65.60 % | 62.26 % | 63.93 % |
| 文中方法 | 71.16 % | 69.30 % | 70.25 % |

3 结束语

提出基于时空兴趣点的人体动作识别方法在公开

运动数据集 KTH 上取得较好的识别性能。实验结果表明 Harris-Laplace 时空兴趣点的 3D-SIFT 描述子特征具有很好的特征描述和类别区分能力,利用 Bag-of-feature 构建词袋,改进的 K-Means 聚类和支持向量机 (SVM) 分类器进行动作分类识别的有效性。相比于现在比较热门的深度学习技术应用到动作识别领域上对实验环境和 GPU 硬件的高需求,比如网络训练阶段,在用梯度反向传播算法计算网络各层参数的过程中,需要保存计算过程中得到的权值 (weight) 和偏置 (bias) 的偏导数以及部分网络层的中间输出结果,而这些数据都需要存放在 GPU 有限的显存上,如果在训练过程中采用批量 (batch) 样本训练策略,那么这些显存的占用量就会成倍增长。提出的方法可以在任何一个操作系统的 MATLAB 平台上实验,对硬件的要求不高。但是相对于真实生活场景中各种复杂的人与人或是人与物之间的交互动作,以及现实场景的复杂多变而言,实验的 KTH 数据集就比较的局限单一,只是包含单个个体在 4 个不同场景的 6 类常见的体育动作。因此,要实现技术商用化的目的,还要在后续的研究工作中对算法等技术进行更加深入的改进研究,同时也要对数据集的多样性进行扩充。

参考文献:

- [1] 张博宇,刘家锋,唐降龙. 一种基于时空兴趣点的人体动作识别方法[J]. 自动化技术与应用, 2009, 28(10): 75-78.
- [2] 王见,陈义,邓帅. 基于改进 SVM 分类器的动作识别方法[J]. 重庆大学学报, 2016, 39(1): 12-16.
- [3] Watanabe T, Tanaka T. Vein authentication using color information and image matching with high performance on natural light [C]. Proceedings of International Joint Conference. Fukuoka: Fukuoka International Congress Center, 2009: 3625-3629.
- [4] Yang M Y, Cao Y, McDonald J. Fusion of camera images and laser scans for wide baseline 3D scene alignment in urban environments[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66: 52-61.
- [5] An Hangxing, Meng Lingjun, Zhao Lin, et al. Long-distance transmission and high-speed and real-time storage technology of image data[J]. Video Engineering, 2013, 37(3): 175-178.
- [6] Cao Y, McDonald J. Improved feature extraction and matching in urban environments based on 3D view point normalization[J]. Computer Vision and Image Understanding, 2012, 116: 86-101.
- [7] W He. Recognition of human activities using a multiclass relevance vector machine [J]. Optical Engineering SPIE, 2012.
- [8] Anoop Cherian, Piotr Koniusz, Stephen Gould. Higher-order Pooling of CNN Features via Kernel Linearization for Action Recognition [C]. IEEE Winter Conference on Applications of Computer Vision, 2017: 130-137.
- [9] Laptev I. On Space-Time Interest Points[J]. International Journal of Computer Vision, 2005, 64(2/3): 432-439.
- [10] Park S, Aggarwal J K. A Hierarchical Bayesian Network for Event Recognition of Human Actions and Interactions[J]. Multimedia Systems, 2004, 10(2): 164-179.
- [11] 丁松涛,曲仕茹. 基于改进时空兴趣点检测的人体行为识别算法[J]. 西北工业大学学报, 2016, 34(5): 886-892.
- [12] Guan T, Wang C. Registration based on scene recognition and natural features tracking techniques for widearea augmented reality systems [J]. IEEE Transaction on Multimedia, 2009, 11(8): 1393-1406.
- [13] C Schuldt, I Laptev, B Caputo. Recognizing Human Actions: A Local SVM Approach [C]. Proceedings of the 17th International Conference on Pattern Recognition. ICPR, 2004.
- [14] Sushideep Narayana. Action Recognition from video [EB/OL]. https://github.com/Sushirdeep/Action-Recognition-from-Videos/blob/master/ProjectReport/ActionRecognitionReport_Sushirdeep.pdf.
- [15] Arthur D, Vassilvitskii S. k-means ++: the advantages of careful seeding [C]. Eighteenth Acm-Siam Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2007: 1027-1035.
- [16] KTH Database [EB/OL]. <http://www.nada.kth.se/cvap/actions/>.
- [17] 付朝霞,王黎明. 基于时空兴趣点的人体行为识别[J]. 微电子学与计算机, 2013, 30(8): 28-35.
- [18] Dufournaud Y, Schmid C, Horaud R. Matching Images with Different Resolutions [C]. The IEEE Con-

- ferences on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, 2000.
- [19] Mokhtarian F, Suomela R. Curvature Scale Space Based Image Corner Detection [C]. European Signal Processing Conference, Island of Rhodes, Greece, 1998.
- [20] Lindeberg T. Feature Detection with Automatic Scale Selection [J]. International Journal of Computer Vision, 1998, 30(2): 79–116.
- [21] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition [C]. DBLP, 2007: 357–360.
- [22] Hartigan J A, Wong M A. A K-means clustering algorithm [J]. Applied Statistics, 1979, 28(1): 100–108.
- [23] 梁宇宏, 张欣. 对遗传算法轮盘赌选择方式的改进 [J]. Information Technology, 2009, 33(12): 127–129.
- [24] 夏桂梅, 曾建潮. 一种基于轮盘赌选择遗传算法的随机微立群算法 [J]. 计算机工程与科学, 2007, 29(6): 51–54.
- [25] 蔡军, 邹鹏, 沈弼龙, 等. 改进轮盘赌策略的反馈式模糊测试方法 [J]. 四川大学学报, 2016, 48(2): 133–137.
- [26] 万士宁. 基于卷积神经网络的人脸识别研究与实现 [D]. 成都: 电子科技大学, 2016.

Human Action Recognition based on Spatio-Temporal Interest Point

CHEN Sheng-di, HE Bing-qian, CHEN Si-yu, LIU Ji-yuan

(College of Computer Science and Technology, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Human action recognition are increasingly attracting much attention from computer vision and pattern recognition researchers. This paper presents a method based on Harris-Laplace algorithm combined with 3D-SIFT descriptor, and a Bag-of-feature approach is used to represent videos. The Harris-Laplace algorithm is used to extract the spatial and temporal interest points. The 3D-SIFT descriptor can better describe the essential characteristics of the video sequence and it is more effective than the traditional descriptor. The K-Means approach is used for clustering. Finally, the support vector machine (SVM) is used as the classifier for human action recognition. One-vs-one and one-vs-rest classification strategies are used and the comparison is made. The experiment on the public database KTH proves the effectiveness and robustness of this method.

Keywords: computer applications technology; image processing and graphics; spatial and temporal interest points; action recognition; harris-laplace; 3D-SIFT; feature extraction